

KOLUMNENTITEL: Hausaufgabenqualität im Französischunterricht

Hausaufgabenqualität im Französischunterricht aus Sicht von Schülern, Lehrkräften und
Experten und die Entwicklung von Leistung, Hausaufgabensorgfalt und Bewertung der
Hausaufgaben

Homework quality in French as a foreign language as perceived by students, teachers, and
experts, and the development of achievement, homework effort and perceived homework
value

Inge Schnyder,

Universität Freiburg (Schweiz) & Pädagogische Hochschule Freiburg (Schweiz)

Alois Niggli, Pädagogische Hochschule Freiburg (Schweiz)

Ulrich Trautwein, Universität Tübingen

Zeitschrift für Pädagogische Psychologie, 22 (3-4), 233-246.

Autorenhinweis: Die vorliegende Studie wurde mit Mitteln des Schweizerischen
Nationalfonds zur Förderung der wissenschaftlichen Forschung (Projekt Nr. 13DPD3-
114176/1) unterstützt.

Zusammenfassung

In dem vorliegenden Beitrag wird eine konzeptuelle Fundierung von Hausaufgabenqualität erarbeitet und zwischen verschiedenen Perspektiven (Schüler, Lehrer, Beobachter) bei der Beurteilung von Hausaufgaben unterschieden. Im empirischen Teil der Studie wurden 1382 Schülerinnen und Schüler aus drei Schweizer Kantonen zu Beginn und am Ende der achten Jahrgangsstufe mit Hilfe von Leistungstests und Fragebögen untersucht. Zudem wurden die Lehrpersonen der Schülerinnen und Schüler befragt. Insgesamt 59 Lehrpersonen machten Angaben über ihre Hausaufgabenpraxis und gaben eine Bewertung ihrer Hausaufgaben; die Lehrkräfte protokollierten außerdem über mehrere Tage die von ihnen erteilten Hausaufgaben, die anschließend von Experten bewertet wurden. Die von Schülerinnen und Schülern sowie von Experten beurteilte Qualität stand hingegen in einem positiven Zusammenhang mit der Entwicklung von Schulleistung, Hausaufgaben Sorgfalt und der Bewertung der Nützlichkeit der Hausaufgaben. Die von den Lehrpersonen eingeschätzte Qualität der Hausaufgaben stand dagegen in keiner statistisch signifikanten Beziehung mit den Outcomevariablen.

Schlagwörter: Hausaufgaben, Hausaufgabenqualität, Fremdsprachen, Schulleistung

Abstract

This article presents a theoretical conceptualization of homework quality and distinguishes three perspectives on the evaluation of homework quality (students, teachers, observers). In the empirical part of the study, 1382 students from three Swiss cantons were administered achievement tests and questionnaires at the beginning and end of grade 8. The students' teachers were also surveyed. A total of 59 teachers reported on their homework practices and evaluated their homework assignments; moreover, they kept a record of the homework assignments set over several days, which were then evaluated by experts. Students' and experts' ratings of homework quality were positively related to the development of school achievement, homework effort, and perceived homework value. No statistically significant predictive effect was found for the teacher-perceived homework quality.

Keywords: Homework, homework quality, second language acquisition, school achievement

Hausaufgabenqualität im Französischunterricht aus Sicht von Schülern, Lehrkräften und Experten und die Entwicklung von Leistung, Hausaufgabensorgfalt und Bewertung der Hausaufgaben

1. Einleitung

Über die Qualität von Hausaufgaben wird seit Jahrzehnten heftig diskutiert. Kritiker bemängeln, dass Hausaufgaben didaktisch-methodisch wenig vorbereitet würden, vor allem aus mechanisch-reproduktiven und monotonen Arbeiten bestünden, die selten im Unterricht fortgeführt würden, kognitiv wenig herausfordernd seien und nur selten an die Bedürfnisse der Schülerinnen und Schüler adaptiert würden (vgl. Bennett & Kalish, 2006; Petersen, Reinert & Stephan, 1990; Rossbach, 1995; Schwemmer, 1980). Dass diese Kritik an Hausaufgaben nicht rein „akademischer Natur“ ist, werden vermutlich Millionen von Eltern bestätigen, die von ihren Kindern in oftmals emotional aufgeladene Diskussionen über den Sinn und Unsinn der Hausaufgaben im Allgemeinen bzw. der Hausaufgaben des jeweiligen Tages verwickelt werden. Qualitätsaspekte sind denn auch Bestandteil differenzierter und empirisch geprüfter Modellvorstellungen zur Hausaufgabenvergabe (Trautwein, Lüdtke, Schnyder & Niggli, 2006).

Angesichts der unverkennbaren Relevanz der Hausaufgabenqualität mag verwundern, wie spärlich die empirische Forschungslage zu diesem Thema noch immer ausfällt (vgl. Trautwein & Köller, 2003; Trautwein & Lüdtke, in press). Dies gilt speziell im Kontrast zur allgemeinen Unterrichtsforschung, in der die Relevanz der Unterrichtsqualität inzwischen programmatisch ist (vgl. Kunter & Baumert, 2006). In der vorliegenden Studie wollen wir einen Beitrag zur Überwindung dieses Forschungsdesiderats leisten, indem wir mit Hilfe einer großen Längsschnittstudie den Zusammenhang zwischen der Hausaufgabenqualität und der Entwicklung von Schulleistung, der Hausaufgabenerledigung und der positiven Bewertung der Hausaufgaben untersuchen.

2. Hausaufgabenforschung: Von der Quantität zur Qualität

Quantitative Aspekte und damit das Interesse an der häuslichen Lernzeit überwogen in der Hausaufgabenforschung bis in die 1990er Jahre, bedingt auch durch die Unterrichtsforschung, die bis Ende der 1980er Jahre hier den Schwerpunkt setzte (Helmke, Hornstein & Terhart, 2000). In einer Metaanalyse jüngeren Datums fanden Cooper, Robinson und Patall (2006) in der Mehrzahl der berücksichtigten Arbeiten einen Zusammenhang zwischen dem Zeitaufwand bei Hausaufgaben und der Schulleistung. Die in die Metaanalyse eingehenden Arbeiten haben jedoch meist auf eine Unterscheidung von Effekten der Hausaufgabenvergabe und Hausaufgabenerledigung verzichtet, was die Interpretation der Befunde erschwert (vgl. Trautwein, 2007). Im deutschen Sprachraum dominierten lange Zeit Arbeiten, die keine Hinweise auf einen substanziellen Zusammenhang zwischen Hausaufgabenmenge und Leistung erbrachten (z.B. Wittmann, 1964) und als Basis für eine umfassende Kritik an Hausaufgaben (z.B. Nilshon, 2001; Schwemmer, 1980) dienten. In letzter Zeit konnte jedoch eine Reihe von Arbeiten einen Zusammenhang zwischen der Häufigkeit der Hausaufgabenvergabe (z.B. Lipowsky, Rakoczy, Klieme, Reusser & Pauli, 2004; Trautwein, 2007) bzw. der Sorgfalt bei der Hausaufgabenerledigung (z.B. Haag & Mischo, 2002; Schnyder, Niggli, Cathomas, Trautwein & Lüdtke, 2006) auf der einen Seite und der Leistungsentwicklung auf der anderen Seite belegen.

Auch zur Qualität bzw. zur Beschaffenheit oder Eigenart von Hausaufgaben liegen Studien vor, von denen aber nur eine Minderheit versuchte, den Zusammenhang zwischen der Qualität der Hausaufgaben und pädagogisch erwünschten Outcomes wie der Steigerung der Leistung, des Interesses, der Motivation oder der Entwicklung von Sorgfalt bei der Hausaufgabenerledigung empirisch zu ermitteln. Petersen et al. (1990) beispielsweise konnten auf Grund einer Lehrerbefragung zeigen, dass die Hälfte der Hausaufgaben dem Üben (52%) dienten und die restlichen Aufgaben als Ziel das Zusammenfassen, Übertragen, Vorbereiten und Wiederholen verfolgten, wobei Wiederholen nach dem Üben den größten Teil der

Aufgaben ausmachte (21%). Schwemmer (1980), der sich größtenteils auf Studien aus den 1960er Jahren bezog, bezeichnete Hausaufgaben mit Übungscharakter sowohl in Hinblick auf die Leistungssteigerung als auch zur Erreichung erzieherischer Ziele als wertlos.

Lipowsky et al. (2004) gingen der Frage nach, wie sich die aus Schülersicht wahrgenommene Hausaufgabenpraxis im Mathematikunterricht auf die spätere mathematische Leistung auswirkt. Die Ergebnisse weisen darauf hin, dass eine positivere Leistungsentwicklung in denjenigen Klassen zu beobachten war, in denen die Lehrperson Hausaufgaben stellte, die die Schülerinnen und Schüler über etwas Neues nachdenken ließen. Die Autoren schlossen daraus, dass die Frage nach der Qualität der Hausaufgaben wichtiger ist als die Frage, ob überhaupt Hausaufgaben gestellt werden. Die Vernachlässigung des didaktischen Aspektes könnte auch einer der Gründe sein, vermuteten die Autoren, warum es widersprüchliche Resultate zur Effektivität von Hausaufgaben gibt, und sie mahnten eine verstärkte Auseinandersetzung der Hausaufgabenforschung mit didaktischen Fragestellungen an.

Hinweise auf einen Zusammenhang zwischen Hausaufgabenqualität und der Erledigung von Hausaufgaben finden sich auch in mehreren Studien der Arbeitsgruppe um Trautwein (vgl. Trautwein & Köller, 2003; Trautwein et al., 2006; Trautwein & Lüdtke, 2007). Nach dem von Trautwein und Kollegen postulierten Modell hat die Hausaufgabenqualität einen direkten Effekt auf die Hausaufgabenmotivation der Schülerinnen und Schüler sowie einen teilweise über die Hausaufgabenmotivation medierten Effekt auf die Hausaufgabenerledigung und die Schulleistung. In empirischen Studien zur Prüfung des Modells wurde die Hausaufgabenqualität über einen Schülerfragebogen erfasst und auf Klassenebene, Individualebene und intraindividuelle Ebene in Beziehung zu der von den Schülerinnen und Schülern berichteten Sorgfalt bei der Hausaufgabenerledigung gesetzt. Auf Klassenebene zeigte sich, dass Schülerinnen und Schüler aus Klassen, in denen die Hausaufgabenqualität insgesamt als hoch beschrieben wurde, auch über ein höheres

Hausaufgabenengagement berichteten als Schülerinnen und Schüler aus Klassen, in denen die Hausaufgabenqualität insgesamt kritischer eingeschätzt wurde. Ein vergleichbares Befundmuster ergab sich auf Individualebene: Wer eine höhere Hausaufgabenqualität wahrnahm als die Klassenkameraden, berichtete über ein höheres Hausaufgabenengagement. Schließlich fand sich auf intraindividueller Ebene, dass die meisten Schülerinnen und Schüler für solche Fächer über mehr Sorgfalt bei den Hausaufgaben berichteten, bei denen sie eine im Vergleich mit den anderen Fächern hohe Hausaufgabenqualität wahrnahmen.

Die genannten Studien zeichnen ein überraschend konsistentes Bild von der Bedeutung der Hausaufgabenqualität. Eine zentrale Einschränkung besteht jedoch darin, dass die Hausaufgabenqualität ausschließlich auf dem Selbstbericht von Schülerinnen und Schülern beruht, die nach eher globalen Einschätzungen der Hausaufgabenqualität gefragt wurden. In der vorliegenden Studie soll diese Einschränkung überwunden werden, indem auf der Basis didaktisch-inhaltlicher Überlegungen eine Beschreibung von Hausaufgabenqualität entwickelt wird, die einer empirischen Operationalisierung im Rahmen eines multiperspektivischen Verfahrens zugänglich ist.

3. Theoretische Verortung der Hausaufgabenqualität

Empirische Befunde über unterrichtliche Qualitätsmerkmale werden breit diskutiert (Ditton, 2006; Heid, 2000; Helmke et al., 2000; Lipowsky, 2006). Hinsichtlich wichtiger Kernelemente scheint inzwischen ein gewisser Konsens feststellbar zu sein (Helmke, 2003; Shuell, 1996). Eine vergleichbare Entwicklung hat für die Vergabe von Hausaufgaben bislang nicht stattgefunden (vgl. Trautwein, Niggli, Schnyder & Lüdtke, in press). Dies mag auch damit zusammenhängen, dass „eine systematische Analyse der Qualität von Aufgaben zu den eher vernachlässigten Forschungsbereichen“ gehört (Blömeke, Risse, Müller, Eichler & Schulz, 2006, S. 331). Wenn der Inhalt von Hausaufgaben überhaupt zum Thema gemacht wird, dann wird ihre Qualität primär bemängelt. So wird etwa argumentiert, Hausaufgaben seien häufig rein repetitiv und monoton (Petersen et al., 1990; Rossbach, 1995; Schwemmer,

1980). Daraus lassen sich zwei Forderungen an die Qualität von Hausaufgaben ableiten. Die erste betrifft die *kognitiven Anforderungen*, die an die Lernenden gestellt werden und bedeutet, dass Hausaufgaben nicht ausschließlich aus mechanisch-repetitiven Übungen bestehen sollten. Die zweite Forderung meint, dass eine gewisse Variabilität der Hausaufgaben bzw. ein angemessener *Anregungsgehalt* notwendig sei.

Kognitive Anforderungen bei den Hausaufgaben konzeptualisieren wir im Folgenden entlang der beiden Pole *reproduktiv-produktiv*. Die kognitiven Aktivierungspotentiale in diesen beiden Bereichen sind unterschiedlich breit. Jedes Fach enthält beispielsweise Wissensanteile, die man auswendig lernen muss, die sich nicht durch Überlegen bzw. Problemlösen erschließen lassen, zum Beispiel Vokabeln lernen. Von dieser einen reproduktiven Grundform zu unterscheiden ist *produktives*, das heißt *elaborierendes* Üben (Edelmann, 2000). Erst durch diese kognitiv aktivierenden Übungsformen wird der Lerngegenstand vertieft erschlossen (Lipowsky, 2006; Steiner, 1996). Auch in der Sprachlernforschung existieren vergleichbare Auffassungen. Valette und Disick (1972) orientieren sich in ihrer Konzeption der kognitiven Ich-Beteiligung beim Lernen einer Fremdsprache ebenfalls an reproduktiven (Reproduzieren, automatisches Abrufen) und produktiven Stufen (situativ abrufen, eigene Gedanken ausdrücken). Ellis (2005, S. 5) unterscheidet bei seiner Beschreibung des aufgabenorientierten Ansatzes im Fremdsprachenunterricht „*exercises*“ (verstanden als Training von Vokabeln und grammatikalischen Formen) und sogenannte „*tasks*“, die er als elaborierten Umgang mit konkreten Inhalten versteht. Im Anschluss an die PISA-Studien werden neuerdings eher Aufgaben konstruiert, die sich zu bestimmten Kompetenzniveaus in Beziehung setzen lassen (Klieme et al., 2006; Nold & Rossa, 2007). Diese anspruchsvollen Konzepte haben aber erst ansatzweise Einzug in den Schulalltag gefunden.

Bei der Orientierung an den beiden Polen *reproduktiv-produktiv* könnte sich die Relation „je anspruchsvoller, desto besser“ allerdings als problematische Auffassung

entpuppen. Dass im Fremdsprachenunterricht auch einfache Fragen und Aufgaben bedeutsam sind, belegen insbesondere die Befunde der Studie *Deutsch-Englisch-Schülerleistungen-International (DESI)* (Klieme et al., 2006) zum Hörverständnis. Einiges spricht dafür, dass das Lernen anspruchsvollerer fremdsprachlicher Inhalte im schulischen Unterricht beispielsweise auf vorhergehendes Automatisieren des Wortschatzes angewiesen ist (Oser & Baeriswyl, 2001). Eine ausgewogene Variation der kognitiven Anforderungen scheint deshalb für die Hausaufgaben von nicht zu unterschätzender Bedeutung zu sein: Einerseits müssen Hausaufgaben einfach genug sein, damit Schülerinnen und Schüler mit einem vertretbaren Aufwand ein gutes Resultat erreichen. Zugleich führen Lernaufgaben erst dann zu intensiveren Vernetzungen, wenn sie entweder neu oder ausreichend herausfordernd sind (Blömeke et al., 2006). Eine Aufgabenpraxis, die sich an diesen Kriterien orientiert, und das persönliche Anspruchsniveau, das jeder einzelne für sich definieren kann, nicht deutlich verfehlt, scheint sich sowohl auf das Lernen als auch auf das Fachinteresse motivierend auszuwirken (Brophy & Good, 1986).

Eine qualitätsvolle Bearbeitung von Aufgaben kann sich nicht ausschliesslich nur auf kognitive sprachliche Anforderungen stützen. Im Rahmen subjektiver Erfahrungsqualitäten kann zusätzlich auch der *Anregungsgehalt* der Hausaufgaben eingeschätzt werden. Darunter wird die Variation innerhalb einer Reihe von Aufgabenstellungen verstanden. Dies betrifft insbesondere unterschiedliche sprachliche Kontexte und variierte mediale Zugangsweisen. Repetitive Übungen oder das Lernen von Vokabeln beispielsweise können auf monotone Art und Weise oder auch variantenreich konzipiert werden. Für Hausaufgaben im Fremdsprachenunterricht sind deshalb insbesondere auch die damit zusammenhängenden motivationalen und volitionalen Aspekte wichtig, die den Lernerfolg gerade in diesen Fächern maßgeblich beeinflussen können (Dörnyei, 2003). Dabei spielt, wie Roth (1976, S. 332) bereits in geradezu klassischer Weise formuliert hat, die Variation der Lernaufgaben eine zentrale Rolle: „Übungen unter immer wieder neuen Gesichtspunkten, an immer wieder

neuem Material, in immer wieder neuen Zusammenhängen, anderen Anwendungen, unter immer wieder neuen größeren Aufgaben – darin steckt das Geheimnis des Übens.“ Die Perspektive von Roth weist damit über reine kognitive Anforderungen hinaus und umfasst auch die Verwendung unterschiedlicher Materialien sowie variable kontextuelle Bezüge. Auf die Bedeutung der Variabilität von Aufgabenstellungen und Lernumgebungen wurde denn auch in jüngerer Zeit wiederholt hingewiesen (Bönsch, 2000; Preuss, 2006; Weinert & Helmke, 1997). In einer Befragung von Lehrpersonen konnte Zahorik (1996) belegen, dass diese der Variation von Aufgaben und Material tatsächlich einen motivierenden Einfluss zuschreiben.

Kognitive Anforderungen und die Variation der verlangten Aktivitäten sind nicht die einzigen möglichen Kriterien, um die Qualität von Hausaufgaben einzuschätzen. Zu nennen sind beispielsweise Adaptivität, Lernprozessphase oder Sprechakt. Adaptivität bedeutet, Unterschiede bei den Lernenden wahrzunehmen und eine auf die lernenden Individuen und Gruppen bezogene Passung zu einem Vergabeprinzip bei Hausaufgaben zu machen. Mit der Einordnung in den Lernprozess sind Hausaufgaben gemeint, die vorbereitenden oder nachbereitenden Charakter haben. Bei vorbereiteten Hausaufgaben beschäftigen sich die Schülerinnen und Schüler mit Informationen, die in einer der kommenden Lektionen im Unterricht aufgegriffen werden. Nachbereitende Hausaufgaben sind Hausaufgaben, die sich aus dem Unterricht ergeben und der Konsolidierung des behandelten Stoffes dienen. Die Sprechakttheorie (Searle 1969) sieht in sprachlichen Äußerungen nicht einfach die Nennung eines Sachverhalts. Vielmehr tun wir etwas mit unseren Wörtern, wir handeln also mit der Sprache. Eine zentrale Funktion des fremdsprachlichen Unterrichts kann infolgedessen darin bestehen, in der Fremdsprache auch Mitteilungsabsichten zu realisieren und solche zu verstehen. Wir hatten die Hausaufgaben in der vorliegenden Studie zusätzlich auf diese Dimensionen untersucht. Da jedoch kaum Varianz zwischen den Lehrpersonen festgestellt werden konnte, werden diese Aspekte im Weiteren nicht weiter verfolgt (vgl. Schnyder,

2006).

Bei den oben genannten Merkmalen handelt es sich um Aspekte der Beschaffenheit des Unterrichtsgegenstandes „Hausaufgaben“. Kognitive Anforderungen können beispielsweise als reproduktiv oder produktiv kategorisiert werden. Aufgaben können ähnlich gestaltet oder variiert sein. Entsprechende Aussagen beschreiben einen Sachverhalt und sind nicht wertend (Helmke, 2003). Grundsätzlich ist es aber auch möglich, Qualität im Sinne von Exzellenz zu beurteilen. In diesem Fall wird eine normative Aussage zum „allgemeinen Niveau“ eines Objektes gemacht (Terhart, 2002, S. 50). Anwendung findet dann ein Gütemaßstab, der beispielsweise von „ungenügend“ bis „sehr gut“ schwanken kann. Im Zentrum stehen nicht mehr einzelne Merkmale der Hausaufgaben. Abgegeben wird in ein umfassendes, holistisches Urteil, das bei den Urteilern eher auf intuitive Weise zustande kommt und sich an einem Bündel von Kriterien orientiert.

Zusammenfassend lässt sich konstatieren, dass – obwohl es an genauen Vorstellungen fehlt, wie Hausaufgaben beschaffen sein sollen (und es insbesondere auch einen Mangel an fachspezifischen Überlegungen gibt) – es doch so etwas wie einen Konsens darüber zu geben scheint, was wichtige Qualitätsmerkmale von Hausaufgaben sind. Zum einen sind hierbei insbesondere die kognitiven Anforderungen zu nennen, die oftmals eher zu niedrig gewählt werden dürften, und zum anderen der motivationale Anreizcharakter der Hausaufgaben, der unter anderem durch kontextuelle Variation und auch die erwähnte kognitive Involvierung der Schülerinnen und Schüler gesteigert werden kann. Schließlich ist prüfenswert, ob auch ein holistisches Qualitätsurteil, das sich an globalen Gütekriterien bzw. Standards orientiert, prädiktive Aussagekraft besitzen kann.

4. Hausaufgabenqualität aus unterschiedlichen Perspektiven

Zur Erfassung der Unterrichtsqualität bieten sich vor allem drei Perspektiven an: die der Schülerinnen und Schüler, die der Lehrpersonen und die von Expertinnen und Experten als externe Beobachter. Die letzte hier erwähnte Perspektive ist sehr aufwendig zu erheben und

kommt deshalb eher selten zum Zug; manchmal findet sie in großen Studien Verwendung (z.B. Klieme et al., 2006; Prenzel et al., 2002).

Die einzelnen Perspektiven haben jeweils spezifische Stärken und Schwächen: So können Lehrpersonen und Schülerinnen bzw. Schüler auf einen gemeinsamen Erfahrungshintergrund und damit auch auf eine gegenseitige Vertrautheit rekurrieren, die externen Beobachtern vollständig fehlt. Das gemeinsam Erlebte kann jedoch auch der Grund für ein bestimmtes Verhalten sein, das von den Involvierten nicht mehr erkennbar ist. Schülerinnen und Schüler wiederum teilen mit den Beobachtern die Wahrnehmungsbedingung. Dabei beziehen sich sowohl Schülerinnen und Schüler als auch Beobachterinnen und Beobachter bei ihrer Einschätzung auf das Tun einer Drittperson, der Lehrperson; diese hingegen beurteilt ihre eigene Praxis. Die Ursache von Verhalten wird dadurch wahrscheinlich unterschiedlich attribuiert (vgl. Actor-Observer-Bias von Jones & Nisbett, 1971). In der Literatur wird bei den Lehrpersonen denn auch eine selbstdienliche Verzerrung (Nisbett & Ross, 1980) im Sinn einer Tendenz zur positiven Selbstdarstellung vermutet, ein Phänomen, das in einer Untersuchung von Kunter und Baumert (2006) jedoch nicht bestätigt werden konnte. Die Tatsache, dass Lehrpersonen kaum je Unterricht beobachten oder eine Rückmeldung zu ihrem eigenen Unterricht erhalten (Clausen, 2002), lässt zudem die Vermutung zu, dass sie für dessen Einschätzung wenig Erfahrung haben.

Aber auch die Perspektiven von Lehrpersonen und Beobachtern überschneiden sich teilweise und trennt sie von der Schülerperspektive: Sie verfügen über ein Professionswissen, das eine didaktische Expertise und damit einen Verständnisrahmen mit sich bringt, der den Schülerinnen und Schülern wahrscheinlich fehlt. So können Schülerinnen und Schüler kaum die Makrostruktur von Unterricht nachvollziehen oder anspruchsvolles Üben einschätzen, da dies eine zu hohe Verständnisanforderung verlangen würde (Clausen, 2002).

Clausen (2002) untersuchte die Unterrichtsqualität anhand von Fragebogendaten von Lehrpersonen sowie von Schülerinnen und Schülern und evaluierte zudem

Unterrichtsaufzeichnungen, die von geschulten Ratern in Bezug auf vorgegebene Qualitätskriterien eingeschätzt wurden. Die drei Perspektiven zeigten in der Auswertung nur eine moderate Übereinstimmung, einzig für das Unterrichtsmerkmal „Repetitives Üben“ resultierte eine statistisch signifikante Korrelation.

Dieses Resultat lässt vermuten, dass der Informationsgehalt der einzelnen Perspektiven interessanter ist als eine eventuelle Übereinstimmung zwischen den Perspektiven. Clausen folgert denn auch aus seiner Untersuchung, dass „Schüler, Lehrer und externe Beurteiler gleichermaßen ihre Berechtigung als Informationsquelle (haben), die Brauchbarkeit (...) vom Ziel der konkreten Anwendung ab(hängt). Ist die Vorhersage der kognitiven oder psychosozialen Entwicklungskriterien vorrangiges Ziel, so liegt es nahe, Schülerangaben zu erheben. Geht es primär um eine differenzierte neutrale Beschreibung und Bewertung des unterrichtlichen Geschehens, so empfiehlt sich eher der Einsatz außenstehender Beobachter.“ (Clausen, 2002, S.188)

Es gibt wenige Studien, in die außenstehende Beobachter involviert sind, da dies forschungstechnisch mit sehr hohem Aufwand verbunden ist und entsprechende finanzielle Mittel verlangt. Mittels Beobachtung werden jedoch Erkenntnisse über Lehr-Lernprozesse gewonnen (Reusser, 2003), über die sonst bloß Vermutungen angestellt werden können.

In Bezug auf Hausaufgabenqualität gibt es bis heute kaum Daten aus der Perspektive außenstehender Beobachter; Aussagen über Hausaufgaben basierten jeweils auf Fragebogenerhebungen bei Lehrpersonen sowie bei Schülerinnen und Schülern. Die vorliegende Studie soll hier eine Lücke schließen.

5. Ableitung der Fragestellung

Die vorliegende Studie ist Teil eines größeren Forschungsprojektes zur Wirkung von Hausaufgaben im Fach Französisch als Fremdsprache (vgl. Niggli et al., 2007; Schnyder et al., 2006; Trautwein et al., in press). Schülerinnen und Schüler aus drei Schweizer Kantonen wurden im Laufe der achten Jahrgangsstufe unter anderem nach ihrer

Hausaufgabenerledigung (Sorgfalt) und ihrer Einstellung den Hausaufgaben gegenüber (Nützlichkeit) befragt. Des Weiteren gaben sie eine Einschätzung der Qualität der Lernaufgaben, die sie daheim erledigen müssen. Eingesetzt wurde außerdem ein Französisch-Leistungstest (siehe auch Neumann et al., 2007). Die Lehrpersonen wurden mittels Fragebogen zur Qualität ihrer Hausaufgaben befragt; zudem haben sie während drei auf das Jahr verteilten Wochen die Hausaufgaben gesammelt und protokolliert.

Dieser Artikel berichtet Befunde zu zwei zentralen Fragestellungen. Erstens wird analysiert, wie eng die unterschiedlichen Perspektiven zur Hausaufgabenqualität, wie sie die Schüler- und Lehrerberichte sowie die Ratings von Experten darstellen, zusammenhängen. Basierend auf den Befunden von Clausen (2002) bzw. Kunter und Baumert (2006) kann vermutet werden, dass der Zusammenhang nur moderat ausfallen dürfte. Zweitens untersuchen wir, wie eng die Hausaufgabenqualität mit der Entwicklung von Schulleistung, Sorgfalt bei den Hausaufgaben und der Wahrnehmung von Nützlichkeit der Hausaufgaben zusammenhängt. Hier erwarteten wir in Anlehnung an das Modell und die empirischen Befunde von Trautwein et al. (2006) insgesamt einen positiven prädiktiven Effekt der Hausaufgabenqualität.

6. Methode

6.1 Stichprobe

Die hier berichteten Analysen stützen sich auf Daten von 1382 Achtklässlern (52.9% Mädchen) aus 66 deutschsprachigen Klassen aus den Kantonen Freiburg, Wallis und Luzern, die von insgesamt 59 unterschiedlichen Französischlehrpersonen unterrichtet wurden und die alle hier interessierenden Untersuchungsinstrumente bearbeitet hatten.

Die Erhebung fand im Schuljahr 2003/2004 statt. Die Schülerinnen und Schüler haben während dieses Jahres zweimal einen Leistungstest bearbeitet, die Qualität der Hausaufgaben, ihre Sorgfalt bei der Hausaufgabenerledigung sowie die Nützlichkeit der Hausaufgaben eingeschätzt. Die Administration der Untersuchungsinstrumente übernahmen die jeweiligen

Französischlehrpersonen, die hierfür genaue schriftliche Instruktionen erhalten hatten. Die Lehrpersonen haben während desselben Schuljahrs ebenfalls einen Einstellungsfragebogen ausgefüllt; sie haben zudem in drei über das Schuljahr hinweg vorgegebenen Wochen ihre Hausaufgabenvergabe protokolliert (Oktober 2003, Januar und April 2004).

6.2 Instrumente

Hausaufgabenqualität aus Sicht der Lehrpersonen. Die Qualität der eigenen Hausaufgabenvergabe wurde von den Lehrpersonen anhand einer Skala mit sieben Items eingeschätzt (Beispielitem: „Ich denke mir oft interessante Hausaufgaben für meine Schülerinnen und Schüler aus.“). Es stand hierfür eine vierstufige Antwortskala (1 = *stimmt überhaupt nicht*, 4 = *stimmt ganz genau*) zur Verfügung (Cronbachs Alpha = .77).

Hausaufgabenqualität aus Sicht der Schülerschaft. Die Schüler und Schülerinnen schätzten die Qualität der Hausaufgaben ihres Französischlehrers bzw. ihrer Französischlehrerin über ähnliche Items ein, wie sie auch bei den Lehrkräften verwendet wurden. Die Skala (acht Items, Alpha = .84) reichte von 1 (*stimmt überhaupt nicht*) bis 4 (*stimmt ganz genau*) (Beispielitem: „Unser Lehrer/unsere Lehrerin denkt sich oft interessante Hausaufgaben aus“.).

Sorgfalt. Die Sorgfalt (*homework effort*) der Schülerinnen und Schüler wird hier verstanden als das Bestreben, bei der Erledigung der Hausaufgaben den an sie gestellten Erwartungen zu genügen und sich entsprechend anzustrengen. Dieses Verständnis von Sorgfalt orientiert sich an der Arbeit von Trautwein und Köller (2003). Sie wurde zu beiden Messzeitpunkten über sechs Items erfasst (t1: Alpha = .73; t2: Alpha = .78), die ebenfalls auf der vierstufigen Antwortskala eingeschätzt wurden (Beispiel: „Ich tue mein Bestes bei den Französisch-Hausaufgaben.“).

Nützlichkeit der Hausaufgaben. Den Schülerinnen und Schülern wurde zu Anfang und zum Schluss des Schuljahrs die Frage nach der Nützlichkeit der Hausaufgaben gestellt (Beispiel: „Die Französisch-Hausaufgaben bringen mir nichts.“ – umgepolt). Die Skala

umfasste vier Items und wies eine befriedigende interne Konsistenz auf (t1: Alpha = .74; t2: Alpha = .73).

Leistung. Zu Beginn und am Ende des Schuljahres kam ein standardisierter, für alle Schülerinnen und Schüler identischer Französisch-Leistungstest zur Anwendung. Der Test bestand aus insgesamt 62 (t1) und 48 (t2) Aufgaben im Multiple-Choice bzw. offenen Format. Die Skalierung des Testes erfolgte auf der Basis des einparametrischen Raschmodells (siehe ausführlich bei Neumann et al., 2007). Über ein Anker-Item-Design konnten die beiden Tests auf eine gemeinsame Metrik gebracht werden. Die interne Konsistenz war zu beiden Messzeitpunkten hoch ($KR-20 > .90$).

Qualitätsrating durch Beobachterinnen. Alle Lehrpersonen wurden gebeten, die erteilten Hausaufgaben im Laufe des Schuljahrs in drei vorgegebenen Wochen (im Oktober 2003 sowie Januar und April 2004) zu protokollieren. Zu diesem Zwecke stand den Lehrpersonen ein vorgedrucktes Formular zur Verfügung, auf dem sowohl die Aufgaben notiert wurden wie auch die jeweilige Funktion und die Sprachfertigkeit, welche dabei geübt werden sollte. Das der Aufgabe zugrunde liegende Material sollten die Lehrpersonen am Protokoll jeweils anhängen.

Zwei Expertinnen schätzten als externe Beobachterinnen die Hausaufgabenqualität auf Basis der vorliegenden Hausaufgabenprotokolle ein. Es handelte sich dabei um zwei Erziehungswissenschaftlerinnen mit einer fachdidaktischen bzw. allgemein didaktischen Spezialisierung, die beide in der Lehrer- und Lehrerinnenbildung tätig waren¹. Die Hausaufgabenqualität wurde von den Beobachterinnen mit drei unterschiedlichen Instrumenten eingeschätzt. Sie beinhalten theoretische Konstrukte, die sich teilweise überschneiden (Kognitive Anforderungen, Anregungsgehalt, Qualitäts-Gesamturteil). So spielen beim Anregungsgehalt auch die kognitiven Anforderungen eine Rolle. Der

¹ Die Ratings einer dritten Beobachterin wurden aufgrund einer geringeren Übereinstimmung in den Ratings ausgeschlossen. Diese Beobachterin ist ebenfalls Erziehungswissenschaftlerin, kann jedoch keine Spezialisierung in einer didaktischen oder fachdidaktischen Richtung aufweisen.

Anregungsgehalt umfasst darüber hinaus jedoch auch kontextuell und mediale variierte Aufgaben. Das evaluative Gesamturteil betrifft die erteilten Hausaufgaben einer Lehrperson insgesamt. Es schliesst u. a. auch die obgenannten Kriterien mit ein. Dagegen unterscheidet sich die Form der drei Messinstrumente allerdings deutlich.

In Bezug auf die *kognitiven Anforderungen* schätzten die Beobachterinnen den prozentualen Anteil an produktiven Hausaufgaben auf einer Skala von 1 bis 5 ein: 1 = praktisch nur reproduktive Aufgaben, 2 = einzelne produktive Aufgaben, 3 = ein guter Teil (ungefähr 30%) produktiver Aufgaben, 4 = ein großer Teil (50% oder mehr) produktiver Aufgaben, 5 = praktisch alle Aufgaben sind produktiv. Die Interraterkorrelation für diese Einschätzung betrug $r = .67$, was als Hinweis auf eine befriedigende Übereinstimmung gewertet werden kann (vgl. Wirtz & Caspar, 2002). Die beiden Ratings wurden deshalb gemittelt.

Der *Anregungsgehalt* der Hausaufgaben wurde mithilfe von vier Ratings bestimmt: „Die Lehrperson gestaltet manchmal bestehende Übungsaufgaben um oder gestaltet eigene (interessant).“, „Manchmal gibt diese Lehrkraft einfache, manchmal anspruchsvolle Hausaufgaben.“, „Bei dieser Lehrkraft gibt es immer wieder Hausaufgaben, die die Schülerinnen und Schüler zum Denken auffordern (Phantasie, Meinung).“ sowie „Insgesamt sind die Hausaufgaben bei dieser Lehrkraft sehr abwechslungsreich.“ Bei den vier Ratings lagen die Interraterkorrelation zwischen .59 und .76, was ebenfalls auf eine hinreichende Übereinstimmung weist. Eine aus den vier gemittelten Ratings gebildete Skala wies eine interne Konsistenz von Alpha = .95 auf.

Die Beobachterinnen fällten zudem ein *Qualitäts-Gesamturteil*, das teilweise auf den oben genannten Kriterien beruht: Die Hausaufgaben sind ungenügend (= 1), genügend (= 2), gut (= 3), sehr gut (= 4). Für die Kategorien *ungenügend* und *sehr gut* lag eine Beschreibung vor: „Lehrpersonen, die die Hausaufgaben konsequent nur aus dem Lehrmittel beziehen und diese nicht offensichtlich eine Auswahl für die Schülerinnen und Schüler sind, werden von

den Beobachterinnen grundsätzlich als ungenügend (1) bewertet. Diese Regel soll aber nicht streng angewendet werden; werden die Hausaufgaben trotzdem als genügend empfunden, soll dies so deklariert werden.“ Und „Als *sehr gut* sollen die Hausaufgaben bezeichnet werden, die ausgewogen zwischen Lehrmittel und anderen Aufgaben wechseln, die die Schülerinnen und Schüler ab und zu auch fordern und anregen und die ihnen sinnvoll erscheinen, weil sie darin einen Nutzen für sich erkennen können, etc. Auch wenn viele Hausaufgaben aus dem Lehrmittel stammen, kann eine 4 erteilt werden. Es ist wichtig, mit 4 nicht ideale Hausaufgaben suchen zu wollen, sondern im Bereich des Machbaren sehr gute Hausaufgaben zu honorieren.“ Die Interkorrelation der Raterinnen betrug für das Gesamturteil $r = .84$.

Kontrollvariablen. Als Kontrollvariable wurde vorerst das *Kursniveau* in die Berechnungen mit einbezogen. Im Kanton Wallis werden bis Ende der achten Klasse zwei Leistungsniveaus geführt, während es in Freiburg und Luzern drei sind. In den Analysen wurde zwischen Grundansprüchen (Realschule bzw. Niveau II) und erweiterten Ansprüchen (Sekundarschule/Progymnasium bzw. Niveau I) unterschieden. Da sich in früheren Analysen auch Unterschiede im *Geschlecht* und aufgrund der *Kantonszugehörigkeit* eingestellt hatten (vgl. Neumann, 2007), wurden im Weiteren auch diese Variablen als Kontrollgrößen in die Analysen mit einbezogen.

6.3 Statistisches Vorgehen

Mit den Analysen wurden zwei Ziele verfolgt: Erstens wollten wir klären, wie eng die unterschiedlichen Operationalisierungen von Hausaufgabenqualität zusammenhängen. Hierfür wurde die Enge des Zusammenhangs der auf Klassenebene aggregierten Aussagen der Schülerinnen und Schüler, der Selbstberichte der Lehrpersonen sowie der drei Qualitätsurteile der Beobachterinnen durch Korrelationsanalysen geprüft. Bei diesen Analysen handelt es sich demnach um Analysen auf Klassenebene.

Zweitens untersuchten wir, wie eng die Hausaufgabenqualität mit der Entwicklung von Schulleistung, Sorgfalt bei den Hausaufgaben und der Wahrnehmung von Nützlichkeit

der Hausaufgaben zusammenhängt. Zur Berechnung der Mehrebenenanalysen wurde das Computerprogramm HLM 6 (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2004) verwendet. Bei der Mehrebenenanalyse handelt es sich um eine spezielle Form der Regressionsanalyse, die es erlaubt, simultan Prädiktoren auf mehreren Ebenen zu berücksichtigen. Alle berichteten Modelle sind sogenannte *random intercept* Modelle, die über das *restricted maximum likelihood* Verfahren geschätzt wurden.

Die gängigen Softwarepakete für Mehrebenenanalysen wie das von uns verwendete Programm HLM 6 bieten allein unstandardisierte Regressionskoeffizienten in den Ergebnisausdrucken an, was die Interpretation der Regressionsgewichte aufgrund der oftmals arbiträren Metriken der Prädiktoren und Kriterien erschwert. Um die spätere Interpretation zu vereinfachen, wurden in der vorliegenden Arbeit alle metrischen Individual- und Klassenebene-Variablen *z*-standardisiert ($M = 0$; $SD = 1$). Die entsprechenden Koeffizienten können daher annähernd im Sinne standardisierter Regressionskoeffizienten interpretiert werden.

Für die Interpretation der Effekte der Hausaufgabenqualität auf Klassenebene (Qualitätsurteile der Beobachterinnen, der Lehrpersonen sowie der Schüler und Schülerinnen) haben wir das von Tymms (2004; für eine Anwendung vgl. Trautwein, Gerlach & Lüdtke, in press) vorgestellte Effektstärkenmaß für kontinuierliche Level-2-Variablen in Mehrebenenanalysen herangezogen. Es ist an Cohens *d*-Wert-Maß angelehnt und berechnet sich nach der folgenden Formel:

$$\Delta = 2 \times B \times SD_{\text{predictor}} / \sigma_e,$$

bei der B das unstandardisierte Regressionsgewicht aus dem Mehrebenenmodell ist, $SD_{\text{predictor}}$ die Standardabweichung der Prädiktorvariable auf der Klassenebene und σ_e die Residualstandardabweichung der Outcomevariable auf der Individualebene. Dies soll an einem kleinen Beispiel demonstriert werden. Gegeben sind ein Regressionskoeffizient auf Klassenebene von $B = .30$, eine Standardabweichung der Prädiktorvariable von $SD = 0.35$ und

eine Residualstandardabweichung von $\sigma_e = 0.80$, es resultiert folgende Gleichung:

$$\Delta = 2 \times 0.30 \times 0.35 / 0.80 = 0.26.$$

Diese Effektstärke von $\Delta = 0.26$ würde bedeuten, dass sich der Unterschied in der abhängigen Variable zwischen zwei Klassen, die sich in der Prädiktorvariable um zwei Standardabweichungen unterscheiden, auf 0.26 belaufen würde. Nach Cohen's (1988) Konventionen würde dies einen kleinen Effekt darstellen. Bei der Interpretation der Effekte ist zu beachten, dass die Effektgrößen in der nicht-experimentellen Forschung zur Unterrichtsqualität meist eher gering ausfallen (vgl. Brophy & Good, 1986) und die vorliegende Studie mit ihrer Prüfung von Hausaufgabeneffekten Neuland betritt. Wir bezeichnen deshalb auch kleine Effekte als bedeutsam.

7. Ergebnisse

7.1 Deskriptive Analysen und Skaleninterkorrelation

In Tabelle 1 werden die Mittelwerte der unterschiedlichen Operationalisierungen der Hausaufgabenqualität berichtet.

Tabelle 1 ungefähr hier.

Die Mittelwerte zur Qualitätseinschätzung der Lehrpersonen waren am höchsten ($M = 2.93$) und befanden sich mit dem Mittelwert zur Qualitätseinschätzung der Schüler und Schülerinnen ($M = 2.71$) oberhalb der Skalenmitte; die drei Qualitätsurteile der Beobachterinnen ($M = 1.80$, $M = 2.00$, $M = 2.3$) lagen unter dieser Grenze. Von den drei Einschätzungen der Beobachterinnen fiel die Qualitäts-Gesamtbeurteilung am höchsten aus, die Einschätzung des Anregungsgehalts am niedrigsten.

Tabelle 2 ungefähr hier

In der Tabelle 2 werden die Skaleninterkorrelationen der unterschiedlichen Perspektiven auf die Hausaufgabenqualität berichtet. Auffällig ist, dass die drei Beobachterurteile (Gesamtbeurteilung, Anregungsgehalt, kognitive Anforderungen) sehr eng zusammenhängen. Dieses Ergebnis ist von der Richtung nicht unerwartet, da sich die Operationalisierungen der

drei Konstrukte wie erwähnt teilweise überschneiden, die Höhe der Korrelationen mag aber vielleicht überraschen.

Der Zusammenhang zwischen den unterschiedlichen Perspektiven (Beobachterinnen, Lehrpersonen, Schülerinnen und Schüler) fiel nur moderat aus. Statistisch signifikante Korrelationskoeffizienten fanden sich lediglich zwischen dem Rating des Anregungsniveaus durch die Beobachterinnen auf der einen Seite sowie der Qualitätseinschätzung durch Lehrkräfte sowie Schülerinnen und Schüler auf der anderen Seite. Die Korrelation zwischen Schüler- und Lehrerurteil erwies sich als statistisch nicht signifikant.

7.2 Vorhersage von Sorgfalt, Nützlichkeit und Leistung

Abschließend wurde eine Sequenz von Mehrebenenanalysen durchgeführt, in denen die Leistung, Hausaufgaben Sorgfalt und Bewertung der Nützlichkeit der Hausaufgaben am Ende des Schuljahrs vorhergesagt wurden. Hierfür wurden für jede der drei abhängigen Variablen drei Modelle spezifiziert, in denen jeweils ein Hausaufgaben-Qualitätsmaß aus der Perspektive der Schülerinnen und Schüler, der Lehrpersonen und von externen Expertinnen berücksichtigt wurde. Angesichts der sehr engen Beziehung zwischen den Qualitätsurteilen der Expertinnen verzichteten wir im Folgenden darauf, die Mehrebenenanalysen für alle drei Experten-Qualitätsurteile zu berichten, sondern beschränken uns auf die Gesamtbewertung der Hausaufgaben durch die Raterinnen. Zusätzlich durchgeführte Analysen zeigten allerdings, dass die Befunde für die beiden anderen Qualitätsindikatoren sehr ähnlich ausfielen.

Als Kovariaten wurden auf der Individualebene neben dem Geschlecht auch die Werte der jeweiligen abhängigen Variablen zu t1 eingeführt, so dass – vereinfacht ausgedrückt – der prädiktive Effekt der Qualitätsvariablen auf die Veränderung der jeweiligen abhängigen Variablen geprüft wurde. Auf der Klassenebene wurden Kanton und Schulform (beide in Form von Dummy-Variablen) in allen Modellen statistisch kontrolliert.

In der Tabelle 3 sind die drei Modelle zur Vorhersage der Hausaufgaben Sorgfalt zu t2 dargestellt. Im Modell 1 erwies sich der prädiktive Effekt der Sorgfalt zu t1, der die Stabilität dieser Variable verkörpert, erwartungsgemäß als statistisch signifikant und bedeutsam. Als statistisch signifikant erwies sich außerdem die Zugehörigkeit zum Kanton Freiburg. Zudem fand sich auch ein statistisch signifikanter, aber kleiner prädiktiver Effekt des Globalratings der Hausaufgabenqualität durch die Beobachterinnen ($\Delta = .18$).

Tabelle 3 ungefähr hier

In den nächsten beiden Modellen (Modell 2 und 3) wurden anstatt der Urteile der Beobachterinnen der Selbstbericht der Lehrkräfte über die Qualität ihrer Hausaufgaben (Modell 2) bzw. der Bericht der Schülerinnen und Schüler (Modell 3) eingeführt. Auffällig in diesen beiden Modellen ist der Kontrast in der Vorhersagekraft der beiden Qualitätsindikatoren. Während der Selbstbericht der Lehrkräfte keine Vorhersagekraft für die Hausaufgaben Sorgfalt besaß, fand sich ein statistisch signifikanter Effekt der Schülereinschätzung, dessen Effektstärke sich immerhin auf $\Delta = .34$ belief.

Für die Vorhersage der durch die Schüler zu t2 berichteten Nützlichkeit von Hausaufgaben wurde eine identische Sequenz von drei Modellen gerechnet (vgl. Tabelle 4). Erwartungsgemäß erwies sich der prädiktive Charakter der Nützlichkeit zu t1, der als Stabilität dieser Einschätzung interpretiert werden kann, für alle Modelle als statistisch signifikant. Jungen schätzten die Nützlichkeit der Hausaufgaben im Fach Französisch zum zweiten Messzeitpunkt in allen Modellen geringer ein als die Mädchen. Zudem beurteilten Schülerinnen und Schüler des Kantons Luzern die Nützlichkeit der Hausaufgaben im Fach Französisch in allen Modellen negativer als die Referenzgruppe (Wallis).

In Hinblick auf die Hausaufgabenqualität standen nach Berücksichtigung der Kovariaten wiederum die Schülereinschätzung und der herangezogene holistische Qualitätsindikator der Expertinnen in einem statistisch signifikanten positiven Zusammenhang mit der zum zweiten Messzeitpunkt eingeschätzten Nützlichkeit der

Hausaufgaben. Die beiden Effektstärken erwiesen sich als bedeutsam (Gesamturteil: $\Delta = .31$, Schülereinschätzung der Qualität: $\Delta = .66$).

Tabelle 4 ungefähr hier

Abschließend wurden die entsprechenden Modelle auch für die Leistung zu t2 als abhängige Variable gerechnet. Dabei war der prädiktive Charakter der Leistung zum ersten Messzeitpunkt erwartungsgemäss hoch bedeutsam. Auch die Kantonszugehörigkeit erwies sich durchgängig als statistisch signifikant. Im Vergleich zur Referenzgruppe (Wallis) wiesen Freiburger und Luzerner Schülerinnen und Schüler – nach Kontrolle der Kovariaten – signifikant schlechtere Leistungen im Leistungstest Französisch zum zweiten Messzeitpunkt auf. Zudem zeigten Schülerinnen und Schüler aus den leistungsstärkeren Abteilungen höhere Leistungen als die Schülerinnen und Schüler der Referenzgruppe.

Tabelle 5 ungefähr hier

Von den Hausaufgabenqualitätsindikatoren hatte das globale Qualitäts-Gesamturteil der Beobachterinnen ($\Delta = .33$) einen im Hinblick auf die Schulleistung positiven prädiktiven Effekt. Keine signifikanten Zusammenhänge ergaben sich nach Kontrolle der übrigen Variablen zwischen den Selbsteinschätzungen der Lehrpersonen sowie den Urteilen der Schülerinnen und Schüler und der Leistung im Fach Französisch.

8. Diskussion

Die Qualität von Hausaufgaben zu erfassen ist organisatorisch aufwendig und verläuft aus diesem Grund meist über Fragebogenerhebungen. Einschätzungen werden in der Regel aus der Sicht von Schülerinnen und Schülern, aber auch aus der Perspektive von Lehrpersonen gewonnen. Daraus ergeben sich sowohl Vor- als auch Nachteile (Clausen, 2002), wobei der vermeintliche Vorteil des gemeinsamen schulischen Erfahrungshintergrunds der Lehrpersonen sowie der Schülerinnen und Schüler dann zu einem Nachteil werden kann, wenn zusätzlich kein neutraler Blick von außen auf die Hausaufgaben gerichtet wird. In der vorliegenden Untersuchung haben die Lehrpersonen die Hausaufgaben während drei auf das

Schuljahr verteilten Wochen notiert. Expertinnen haben die Qualität dieser protokollierten Hausaufgaben eingeschätzt. Zusätzlich haben wir die Lehrkräfte selbst sowie ihre Schülerinnen und Schüler um eine Einschätzung der Qualität der Hausaufgaben gebeten.

8.1 Hausaufgabenqualität aus Sicht von Experten und Schülern

Der zentrale Befund unserer Analysen ist der Nachweis, dass Hausaufgaben durch Beobachter zu identifizierende globale Qualitätsmerkmale aufweisen, die es erlauben, die Entwicklung der Sorgfalt bei der Hausaufgabenerledigung, der Einschätzung der Nützlichkeit von Hausaufgaben und schließlich der Leistung vorherzusagen. Lehrpersonen, die vom gängigen Stereotyp reproduktiver Aufgaben abweichen, scheinen entsprechende Entwicklungen begünstigen zu können. Dieses Ergebnis fand sich trotz der vergleichsweise geringen Varianz, die zwischen den Lehrpersonen bestand.

Neben der Einschätzung durch die Expertinnen erwies sich auch das Schülerurteil, nicht aber der Lehrerselbstbericht als prädiktiv. Ein konsistentes Bild lieferten die Befunde zur Entwicklung von Sorgfalt bei der Hausaufgabenerledigung und zur Einschätzung der Nützlichkeit von Hausaufgaben. Mit Ausnahme der Selbsteinschätzung durch die Lehrkräfte konnten für alle Prädiktoren bedeutsame positive statistische Zusammenhänge mit den Outcome-Variablen festgestellt werden. Am deutlichsten war dies bei der Qualitätseinschätzung durch die Schülerinnen und Schüler der Fall. Dies kann zumindest teilweise wohl darauf zurückgeführt werden, dass Schülereinschätzungen in aller Regel die Folge von Generalisierungsprozessen subjektiv wahrgenommener und interpretierter situationsspezifischer Erfahrungen mit den Hausaufgaben sind. Diese Generalisierung betrifft sowohl Merkmale der Vergabe wie der Erledigung von Hausaufgaben.

Die Leistungsentwicklung wurde mithilfe standardisierter Leistungstests untersucht. Hier verfehlte das Schülerurteil zur Hausaufgabenqualität knapp die statistische Signifikanzgrenze. Wie schon Clausen (2002) argumentierte, dürfte die abstrakte Einschätzung von inhaltlichen Konstrukten für Schülerinnen und Schüler relativ schwierig

sein. Dies gilt insbesondere für didaktische Qualitätsmerkmale, die von Experten adäquater zu beurteilen sind. Dass die Einschätzung der Qualität durch Lehrpersonen durchwegs keine prädiktive Aussagekraft hatte, könnte dagegen auf die von Nisbett und Ross (1980) vermuteten selbstdienlichen Verzerrungsprozesse verweisen. Dagegen standen die Qualitätsurteile der Beobachterinnen in einem statistisch signifikanten Zusammenhang mit der Leistungsentwicklung.

Auch für eine Reihe weiterer Variablen, die nicht im Mittelpunkt der vorliegenden Studie standen, fanden sich statistisch signifikante prädiktive Effekte. Jungen schätzten die Nützlichkeit von Hausaufgaben in Französisch signifikant niedriger ein als Mädchen, wohingegen sich in den spezifizierten Modellen keine schlechtere Leistungsentwicklung als bei den Mädchen zeigte. Die Kantonseffekte weisen darauf hin, dass mit der Entfernung zur Sprachgrenze auch das Nützlichkeitsempfinden der Sprache gegenüber abnimmt. Sowohl Freiburg wie auch das Wallis sind zweisprachige Kantone, in denen die andere Sprache eine mehr oder weniger große Rolle spielt. In dem Innerschweizer Kanton Luzern hingegen gibt es für Schülerinnen und Schüler wohl kaum Berührungspunkte mit der französischen Sprache. Auch die signifikant schlechteren Leistungen der Schülerinnen und Schüler aus dem Kanton Luzern im Französisch-Leistungstest könnten auf die Rahmenbedingungen zurückgeführt werden: Schülerinnen und Schüler aus Luzern erhalten insgesamt weniger Französisch-Unterricht (weniger Schuljahre und weniger Unterrichtsstunden) als die Schülerinnen und Schüler aus den beiden anderen Kantonen. Die ebenfalls signifikant schlechteren Leistungen im Französischtest zu t2 der Schülerinnen und Schüler aus dem Kanton Freiburg im Vergleich mit den Schülern aus dem Kanton Wallis lassen sich laut Neumann et al. (2007) vermutlich dadurch erklären, dass die Schülerinnen und Schüler aus dem Wallis in dem entsprechenden Schuljahr intensiver gefördert und zum Lernen angehalten werden, da die Selektion für das Gymnasium nach der 8. Klasse ansteht. Im Kanton Freiburg ist diese Selektion bereits nach dem 6. Schuljahr erfolgt.

8.2 Unterschiedliche Aspekte von Qualität

Die hohen Interkorrelationen der drei Qualitätseinschätzungen (Kognitive Anforderungen, Anregungsgehalt, Gesamturteil) suggerieren, dass die drei Konstrukte möglicherweise deckungsgleich sind. Obwohl sich die Messinstrumente auch in formaler Hinsicht unterscheiden, fand sich eine sehr hohe Interkorrelation, die gegen eine differentielle Vorhersagekraft im Hinblick auf die Outcome-Variablen sprach. Ob die hohen Interkorrelationen dem in der vorliegenden Studie gewählten methodischen Verfahren (z.B. Ratingdimensionen) geschuldet war, von der Auswahl der Raterinnen beeinflusst wurde oder aber der Tatsache geschuldet ist, dass sich die untersuchten Dimensionen tatsächlich nicht trennen lassen, ist natürlich eine Frage, die sich mit den vorliegenden Daten nicht klären lässt. Aus theoretischer Sicht (siehe Einleitungsteil) bestehen durchaus Unterschiede zwischen den drei Aspekten.

Zumindest in der vorliegenden Arbeiten konnten die drei Urteilsaspekte jedoch auf einer einzigen, empirisch nicht weiter differenzierbaren Qualitätsdimension abgebildet werden. Deshalb wurde entschieden, stellvertretend für alle drei Merkmale lediglich das Gesamturteil in die Mehrebenenanalysen aufzunehmen. Dieses Kriterium beruht teilweise auch auf den Merkmalen, die für den Anregungsgehalt und die kognitiven Anforderungen bestimmt worden sind.

Aus Tab. 1 geht hervor, dass reproduktive Aufgaben mit geringer Variation deutlich überwiegen. Dies lässt darauf schliessen, dass Grammatikübungen bzw. das Einprägen sprachlicher Formen, Strukturen und Vokabeln sowohl im Lehrwerk als auch in der Hausaufgabenpraxis nach wie vor dominieren. Produktive, kommunikative Aufgaben dürften dagegen eher als gelegentliche parallele Ergänzungen Berücksichtigung gefunden haben. Die Fokussierung auf nicht kontextualisierte Strukturen bei Lernaufgaben – quasi zum Selbstzweck - wird in den Sprachwissenschaften generell als *focus on formS* thematisiert (vgl. Finkbeiner & Knieriem, 2008, S. 154). Hiervon abzugrenzen ist der *focus on form*, wenn sich

Lernende sprachlichen Strukturen im Rahmen kommunikativer Aufgaben zuwenden können. Gegenüber der Dominanz an reproduktiven Anforderungen könnten parallel konzipierte, kognitiv herausfordernde und durch den kommunikativen Kontext auch als anregend kategorisierte Aufgaben im focus on form eingefangen werden. Aufgrund der vorliegenden Befundlage könnte sich die Qualität der Hausaufgaben im Fremdsprachenunterricht theoretisch deshalb auch an den erwähnten beiden Fokusstrategien orientieren. Die im theoretischen Teil erwähnte Unterscheidung von Ellis (2005) in „exercises“ und „tasks“ weist in eine ähnliche Richtung. Die polaren Beschreibungen der Hausaufgabenqualität, wie sie durch die Dimensionen „reproduktiv-produktiv“ oder beim Gesamturteil bereits verwendet worden sind, haben sich infolgedessen grundsätzlich bewährt und könnten durch entsprechende sprachwissenschaftliche Aspekte allenfalls theoretisch angereichert und in eine verbindende Grunddimension überführt werden. Darin würde auch der Anregungsgehalt eingeschlossen sein, da er aufgrund der doch geringen Varianz zwischen den Lehrkräften wohl auf seine kognitiven Elemente reduziert werden könnte, denn höhere kognitive Leistungen verlangen notwendigerweise auch nach anregenden sprachlichen Kontexten.

Wenn über Inhalte von Hausaufgaben im Fremdsprachenunterricht gesprochen wird, sollte auch über das Verhältnis zwischen Üben/Automatisieren und elaborierten Aufgaben die Rede sein. In der vorliegenden Studie konnte dieses Verhältnis nicht abschließend geklärt werden, weil die Dominanz des häufig kritisierten reproduktiven Übens in der untersuchten Grundpopulation zu ausgeprägt gewesen war (vgl. Schnyder, 2006). Ein zu niedriger Anspruch scheint jedoch aufgrund der vorliegenden Resultate weder der Sorgfalt der Hausaufgabenerledigung noch der Einschätzung der Nützlichkeit zuträglich zu sein (vgl. auch Trautwein et al., in press).

In diesem Zusammenhang wäre ebenfalls prüfenswert, wie die Unterrichtsqualität und die Hausaufgabenqualität zusammenhängen. So ist es denkbar, dass reproduktive Aufgaben (Strukturen automatisieren, Vokabeln lernen) die Hausaufgabenpraxis deutlicher dominieren

könnten als den Unterricht mit der Klasse, wo für kommunikative Aufgaben ein soziales Setting zur Verfügung steht. Einfache reproduktive Grundaufgaben mit Inhalten, die man sich einprägen muss, könnten sogar bewusst den Hausaufgaben anheim gestellt werden, weil man sich damit erhofft, Voraussetzungen für anspruchsvollere Anwendung im Unterricht zu schaffen.

8.3 Grenzen und Perspektiven

Die vorliegende Studie hat Grenzen, auf die abschließend aufmerksam gemacht werden soll. Hierzu gehört die Frage nach der Kausalitätsrichtung. In der vorliegenden Studie haben wir lediglich die prädiktiven Effekte von Hausaufgabenqualität auf die Outcome-Variablen geprüft, nicht aber die ebenso möglichen umgekehrt wirksam werdenden Effekte. So wäre es denkbar, dass sich Lehrpersonen bei den Hausaufgaben auf realisierbare Möglichkeiten beschränken und sich bei den Anforderungen, die sie stellen, dem Leistungs- und Entwicklungsniveau der jeweiligen Klasse anpassen. Schulklassen mit einem höheren Leistungspotential würden dann auch herausforderndere Hausaufgaben erhalten. Niggli, Cathomas, Trautwein und Schnyder (2007) konnten beispielsweise aufzeigen, dass Schulklassen mit Grundansprüchen (vergleichbar mit Hauptschulklassen) anteilmässig mehr Vokabeln zu lernen hatten als Schülerinnen und Schüler in höheren „tracks“, die sich vergleichsweise mehr mit kommunikativen Aktivitäten konfrontiert sahen. Im vorliegenden Design sind diese reziproken Effekte jedoch schwierig zu prüfen, obwohl ihre Bedeutung offensichtlich ist. Wir neigen jedoch der Auffassung zu, dass sich entgegen gesetzte Wirkungen in relativ engen Grenzen bewegen dürften. Die Schulform wurde in der vorliegenden Analyse kontrolliert. Ferner hatte der überwiegende Teil der Lehrpersonen ihre Klassen vor Messbeginn bereits während eines Jahres unterrichtet. Anpassungen an das Niveau der Klasse sollten sich somit bereits eingependelt haben. Somit spricht einiges für die im Längsschnitt erhobenen Effekte in Folgejahr.

Eine weitere Frage betrifft die Validität des Materials, das die Raterinnen vorliegen hatten. Könnten die Lehrpersonen ihre Hausaufgabenpraxis positiver dargestellt haben als ihre gängige Praxis tatsächlich ausfällt? In der Tat sind Verzerrungen nicht ganz auszuschliessen. Eine zusätzliche direkte Unterrichtsbeobachtung durch die Expertinnen hätte wohl weitere, wichtige Informationen liefern können und wäre durchaus wünschenswert gewesen. Insbesondere hätte der kognitiv aktivierende Umgang mit den Hausaufgaben im Gesamtkontext des jeweiligen Unterrichts eingeschätzt werden können. Andererseits hätte die Anwesenheit von Beobachterinnen im Unterricht die Lehrpersonen vermutlich ebenso dazu veranlassen können, ihre Auswahl der Hausaufgaben eher in einem positiven Lichte erscheinen zu lassen.

Zu erwähnen ist auch, dass wir uns auf einige Teilbereiche der Hausaufgabenqualität beschränken mussten. Daneben gibt es weitere Merkmale, die in der Literatur als relevant hervorgehoben werden. Einige dieser Merkmale haben wir erfasst, konnten ihre Effekte aber in der vorliegenden Studie nicht weiter verfolgen, weil die natürlich vorkommende Varianz in diesen Merkmalen nicht ausreichte, um ihre Effektivität zu prüfen. Für eine eingehendere Prüfung wären aus diesen Gründen vermutlich quasi-experimentelle Implementationsdesigns erforderlich. Erst dieses Vorgehen würde es gestatten, relevante Merkmale gezielt zu beeinflussen und eine grössere Varianz zwischen den Lehrkräften hervorzurufen.

Für eine präzisere Analyse der Hausaufgabenqualität generell und im Fremdsprachenunterricht im Speziellen sind somit weitere Forschungsschritte notwendig. Die vorliegenden Einschätzungen der Qualität von Hausaufgaben basieren auf hochinferenten Ratings; eine niedriginferente Auswertung zusätzlicher Daten ist momentan im Gange. Diese Analysen könnten zu zusätzlichen, differenzierteren Erkenntnissen über die Hausaufgabenvergabe und ihre Konsequenzen führen: Mit detaillierteren Angaben über Merkmale der verwendeten Aufgaben wird man den Merkmalen und der Bedeutung von Hausaufgabenqualität vielleicht nochmals einen Schritt näherkommen.

Literaturverzeichnis

- Bennett, S. & Kalish, N. (2006). *The case against homework: How homework is hurting our children and what we can do about it*. New York: Crown Publishers.
- Blömeke, S., Risse, J., Müller, C., Eichler, D. & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. *Unterrichtswissenschaften*, 34, 330–357.
- Bönsch, M. (2000). *Variable Lernwege*. Stuttgart: UTB.
- Brophy, J. & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (S. 328–375). New York: Macmillan.
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive?* Münster: Waxmann.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooper, H., Robinson, J. C. & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research*, 76, 1–62.
- Ditton, H. (2006). Unterrichtsqualität. In K. H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 235–243). Bad Heilbrunn: Klinkhardt.
- Dörnyei, Z. (2003). Attitudes, orientations and motivation in language learning: Advances in theory, research and applications. In Z. Dörnyei (Ed.), *Attitudes orientations and motivation in language learning: Advances in theory, research and applications* (pp. 3–32). Oxford: Blackwell Publishing, Ltd.
- Edelmann, W. (2000). *Lernpsychologie*. Weinheim: Beltz.
- Ellis, R. (2005). *Instructed second language acquisition: A literature review*. Ministry of Education, New Zealand. <http://www.educationcounts.edcentre.govt.nz/publications/downloads/instructed-second-language.pdf>

- Finkbeiner, C. & Knieriem, M. (2008). Aufgabenorientiertes Lernen im Fremdsprachenunterricht: Beispiele zur Förderung kognitiver, metakognitiver und sozioaffektiver Lernprozesse. In J. Thonhauser (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen*, (S. 149 – 168). Münster: Waxmann.
- Haag, L. & Mischo, C. (2002). Saisonarbeiter in der Schule – einem Phänomen auf der Spur. *Zeitschrift für Pädagogische Psychologie*, 16, 109–115.
- Heid, H. (2000). Qualität: Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. In A. Helmke, W. Hornstein & E. Therhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule* (S. 41–54). Weinheim: Beltz (Zeitschrift für Pädagogik, 41. Beiheft).
- Helmke, A. (2003). *Unterrichtsqualität*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Helmke, A., Hornstein, W. & Terhart, E. (2000). *Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule*. Weinheim: Beltz (Zeitschrift für Pädagogik, 41. Beiheft).
- Jones, E. E. & Nisbett, R. E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones (Ed.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R. H., Nold, G., Rolff, H.-G., Schröder, K., Thomé, G. & Willenberg, H. (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI)*. Frankfurt a. M.: Deutsches Institut für Internationale Pädagogische Forschung.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9, 231–251.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der

- Schuler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf* (S. 47–70). Weinheim: Beltz (Zeitschrift für Pädagogik, 51. Beiheft).
- Lipowsky, F., Rakoczy, K., Klieme, E., Reusser, K. & Pauli, C. (2004). Hausaufgabenpraxis im Mathematikunterricht – ein Thema für die Unterrichtsqualitätsforschung? In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderungen als Strategien der Qualitätsverbesserung* (S. 250–266). Münster: Waxmann.
- Lüdtke, O., Trautwein, U., Schnyder, I. & Niggli, A. (2007). Simultane Analysen auf Schüler- und Klassenebene: Eine Demonstration der konfirmatorischen Mehrebenen-Faktorenanalyse zur Analyse von Schülerwahrnehmungen am Beispiel der Hausaufgabenvergabe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39, 1-11.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O. & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus. *Zeitschrift für Erziehungswissenschaft*, 10, 399–420.
- Niggli, A., Cathomas, R., Trautwein, U. & Schnyder, I. (2007). Fachdidaktische Massnahmen im Französischunterricht aus Schülersicht: Der Zusammenhang mit Schulleistung und Motivation. *Schweizerische Zeitschrift für Bildungswissenschaften*. 29, 473 – 504.
- Niggli, A., Trautwein, U., Schnyder, I., Lüdtke, O. & Neumann, M. (2007). Elterliche Unterstützung kann hilfreich sein, aber Einmischung schadet: Familiärer Hintergrund, elterliches Hausaufgabenengagement und Leistungsentwicklung. *Psychologie in Erziehung und Unterricht*, 54, 1-14.
- Nilshon, I. (2001). Hausaufgaben. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 231–238). Weinheim: Beltz.
- Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social*

- judgement*. New York, CA: Prentice-Hall.
- Nold, G. & Rossa, H. (2007). Hörverstehen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 178–196). Weinheim: Beltz.
- Oser, F. & Baeriswyl, F. (2001): Choreographies of Teaching: Bridging Instruction to Learning. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4. ed., pp. 1031–1065). Washington: American Educational Research Association.
- Petersen, J., Reinert, G.-B. & Stephan, E. (1990). *Betrifft: Hausaufgaben: Ein Überblick über die didaktische Diskussion für Elternhaus und Schule*. Frankfurt am Main: Peter Lang.
- Prenzel, M., Seidel, T., Lehrke, M., Rimmel, R., Duit, R., Euler, M., Geiser, H., Hoffmann, L., Müller, C. & Widodo, A. (2002). Lehr-Lern-Prozesse im Physikunterricht – eine Videostudie. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 139–156). Weinheim, Basel: Beltz (Zeitschrift für Pädagogik, 45. Beiheft).
- Preuss, E. (2006). Lehrerkompetenzen zur Gestaltung von Innerer Differenzierung. In W. Plöger (Hrsg.), *Was müssen Lehrerinnen und Lehrer können?* (S. 209–223). Paderborn: Schöningh.
- Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R. & du Toit, M. (2004). *HLM 6 hierarchical linear and nonlinear modeling*. Lincolnwood: Scientific Software International.
- Reusser, K. (2003). *Nutzen von Videoanalysen für die Aus- und Weiterbildung von Lehrpersonen*. Vortrag an der ETH Zürich am 10. Mai 2003.
<http://www.didac.unizh.ch/public/Referate/2003/Videobas_LB03.pdf>
- Rossbach, H.-G. (1995). Hausaufgaben in der Grundschule. Ergebnisse einer empirischen Untersuchung. *Die Deutsche Schule*, 87 (1), 103–112.
- Roth, H. (1976). *Pädagogische Psychologie des Lehrens und Lernens*. Hannover: Schroedel.

- Schwemmer, H. (1980). *Was Hausaufgaben anrichten: Von der Fragwürdigkeit eines durch Jahrhunderte verewigten Tabus in der Hausaufgabenschule unserer Zeit*. Paderborn: Schöningh.
- Schnyder, I. (2006). *Charakteristika der Hausaufgaben im Fach Französisch*. Beitrag an der Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) vom 10.-13. September in München, Maximilians-Universität.
- Schnyder, I., Niggli, A., Cathomas, R., Trautwein, U. & Lüdtke, O. (2006). Wer lange lernt, lernt noch lange nicht viel mehr: Korrelate der Hausaufgabenzeit im Fach Französisch und Effekte auf die Leistungsentwicklung. *Psychologie in Erziehung und Unterricht*, 53, 107–121.
- Shuell, T. J. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Macmillan.
- Searle, J. R. (1969): *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press
- Steiner, G. (1996). Lernverhalten, Lernleistung und Instruktionsmethoden. In *Enzyklopädie der Psychologie* (Bd. 3, S. 278–317). Göttingen: Hogrefe.
- Terhart, E. (2002). *Nach PISA*. Hamburg: Europäische Verlagsanstalt.
- Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, 17, 372–388.
- Trautwein, U., Gerlach, E. & Lüdtke, O. (in press). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of reference effects. *Journal of Educational Psychology*.
- Trautwein, U. & Köller, O. (2003). Was lange währt, wird nicht immer gut: Zur Rolle selbstregulativer Strategien bei der Hausaufgabenerledigung. *Zeitschrift für*

Pädagogische Psychologie, 17, 199–209.

Trautwein, U. & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-student differences and within-student variation. *Journal of Educational Psychology*, 99, 432–444.

Trautwein, U. & Lüdtke, O. (in press). Predicting homework motivation and homework effort in six school subjects: The need to differentiate between the student and class levels. *Learning and Instruction*.

Trautwein, U., Lüdtke, O., Schnyder, I. & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, 98, 438–456.

Trautwein, U., Niggli, A., Schnyder, I. & Lüdtke, O. (in press). Between-teacher differences in homework assignments and the development of students' homework effort, homework emotions, and achievement. *Journal of Educational Psychology*.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.

Valette, R. M. & Disick, R. S. (1972). *Modern Language Performance Objectives and Individualization*. New York: Harcourt.

Weinert, F. E. & Helmke, A. (1997). *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlagsunion.

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität : Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Wittmann, B. (1964). *Vom Sinn und Unsinn der Hausaufgaben – Empirische Untersuchung über ihre Durchführung und ihren Nutzen*. Berlin: Hermann Luchterhand.

Zahorik, J. (1996). Elementary and secondary teachers' reports how they make learning interesting. *Elementary School Journal*, 96, 551–564.

Tabelle 1:

Mittelwerte und Standardabweichungen der Qualitätsindikatoren

Skala/Item	Beispiel/Wortlaut	<i>M</i>	<i>SD</i>
<i>Expertinnenperspektive</i>			
Gesamtbeurteilung	Ungenügend (= 1), genügend (= 2), gut (= 3), sehr gut (= 4)	2.30	.75
Anregungsgehalt	Insgesamt sind die Hausaufgaben bei dieser Lehrkraft sehr abwechslungsreich.	1.80	.63
Kognitive Anforderungen	Anteil an reproduktiven und produktiven Aufgaben	2.00	.81
<i>Schülerperspektive</i>			
Qualität	Unser Lehrer/unsere Lehrerin denkt sich oft interessante Hausaufgaben aus.	2.71	.28
<i>Lehrerperspektive</i>			
Qualität	Ich denke mir oft interessante Hausaufgaben für meine Schülerinnen und Schüler aus.	2.93	.31

Tabelle 2: Korrelationen zwischen den eingesetzten Indikatoren der Hausaufgabenqualität

	(1)	(2)	(3)	(4)
(1) Gesamtbeurteilung (Expertinnen)				
(2) Anregungsgehalt (Expertinnen)	.92***			
(3) Kognitive Anforderungen (Expertinnen)	.81***	.85***		
(4) Qualitätsurteil (Schüler)	.20	.25*	.16	
(5) Qualitätsurteil (Lehrpersonen)	.23	.31*	.20	.23

Anmerkung. * $p < .05$, *** $p < .001$.

Tabelle 3: Vorhersage der Sorgfalt der Hausaufgaben-Erledigung im Fach Französisch zu t2 (Befunde aus Mehrebenenanalysen)

Prädiktoren	Modell 1		Modell 2		Modell 3	
	<i>B</i>	<i>SE(B)</i>	<i>B</i>	<i>SE(B)</i>	<i>B</i>	<i>SE(B)</i>
Kanton FR	-0.14 *	0.07	0.14	0.07	-0.08	0.06
Kanton LU	-0.26	0.15	-0.12	0.07	-0.18	0.13
Track (Sek)	-0.04	0.07	-0.36 **	0.14	-0.05	0.06
Qualitätsurteil Expertinnen	0.08 *	0.03				
Qualitätsurteil Lehrpersonen			0.00	0.04		
Qualitätsurteil Schüler					0.14 ***	0.03
Geschlecht: Jungen	-0.10	0.06	-0.10	0.06	-0.10	0.06
T1 Sorgfalt	0.48 ***	0.04	0.48 ***	0.04	0.47 ***	0.04
Varianzaufklärung						
Level 2	0.60		0.60		0.80	
Level 1	0.22		0.22		0.22	

Anmerkung. * $p < .05$, ** $p < .01$, *** $p < .001$.

Tabelle 4: Vorhersage der wahrgenommenen Nützlichkeit von Französisch-Hausaufgaben zu t2 (Befunde aus Mehrebenenanalysen)

Prädiktoren	Modell 1		Modell 2		Modell 3	
	<i>B</i>	<i>SE(B)</i>	<i>B</i>	<i>SE(B)</i>	<i>B</i>	<i>SE(B)</i>
Kanton FR	-0.10	0.09	-0.07	0.09	-0.00	0.07
Kanton LU	-0.49 **	0.15	-0.59 ***	0.15	-0.33 ***	0.09
Track (Sek)	-0.01	0.10	0.02	0.09	-0.02	0.07
Qualitätsurteil Expertinnen	0.12 **	0.05				
Qualitätsurteil Lehrpersonen			0.07	0.04		
Qualitätsurteil Schüler					0.26 ***	0.03
Geschlecht: Jungen	-0.18 **	0.06	-0.18 **	0.06	-0.18 **	0.06
T1 Nützlichkeit	0.46 ***	0.04	0.46 ***	0.04	0.44 ***	0.04
Varianzaufklärung						
Level 2	0.64		0.59		0.86	
Level 1	0.21		0.21		0.21	

Anmerkung. * $p < .05$, ** $p < .01$, *** $p < .001$.

Tabelle 5: Vorhersage der Französisch-Leistung zu t2 (Befunde aus Mehrebenenanalysen)

Prädiktoren	Modell 1			Modell 2			Modell 3		
	<i>B</i>		<i>SE(B)</i>	<i>B</i>		<i>SE(B)</i>	<i>B</i>		<i>SE(B)</i>
Kanton FR	-0.21	***	0.05	-0.20	***	0.06	-0.19	***	0.06
Kanton LU	-0.57	***	0.14	-0.64	***	0.15	-0.59	***	0.15
Track (Sek)	0.42	***	0.06	0.43	***	0.06	0.42	***	0.06
Qualitätsurteil Expertinnen	0.07	*	0.03						
Qualitätsurteil Lehrpersonen				0.02		0.02			
Qualitätsurteil Schüler							0.05		0.03
Geschlecht: Jungen	-0.03		0.03	-0.03		0.03	-0.03		0.03
T1 Leistung	0.70	***	0.02	0.70	***	0.02	0.70	***	0.02
Varianzaufklärung									
Level 2	0.94			0.94			0.94		
Level 1	0.49			0.49			0.49		

Anmerkung. * $p < .05$, ** $p < .01$, *** $p < .001$.

